NAVAL POSTGRADUATE SCHOOL
Monterey, California

Classification Trees and the
Analysis of Helicopter Vibration Data

by

Harold Larson

September 1997

Approved for public release; distribution is unlimited.

Prepared for: Naval Postgraduate School
                  Monterey, California 93943

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA   93943-5000

Rear Admiral M. J. Evans
Superintendent

Richard Elster
Provost

This report was prepared for and funded by the Naval Postgraduate School.
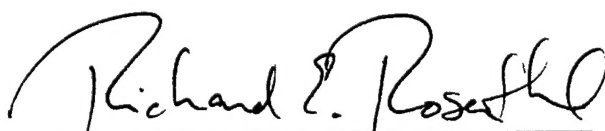
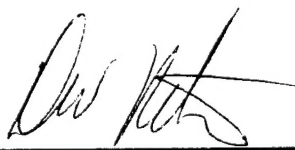Reproduction of all or part of this report is authorized.

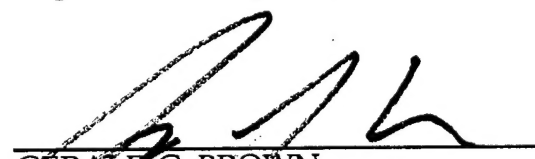This report was prepared by:

HAROLD LARSON
Professor of Operations Research

Reviewed by:

Released by:

RICHARD E. ROSENTHAL
Chairman
Department of Operations Research

DAVID W. NETZER
Associate Provost and Dean of Research

GERALD G. BROWN
Associate Chairman for Research
Department of Operations Research

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1997	3. REPORT TYPE AND DATES COVERED Technical

4. TITLE AND SUBTITLE
Classification Trees and the Analysis of Helicopter Vibration Data

5. FUNDING NUMBERS

6. AUTHOR(S)
Harold Larson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-97-017

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Health and Usage Monitoring Systems are receiving a great deal of interest, in an attempt to increase the safety and operational readiness of helicopters, and to minimize maintenance costs. These systems monitor (and can record) various flight parameters, pilot conversations, engine exhaust debris, metallic chip detector levels in the lubrication system, rotor track and balance, as well as vibration levels at selected locations throughout the airframe and the power drive system. It is hoped that vibration monitoring systems, in particular, may provide useful early warnings of possible mechanical problems in the drive train system, improving the safety and reliability of flight operations.

Vibration levels are currently being observed on two operational SH-60B helicopters and on an H-60 power drive system installed in the Helicopter Transmission Test Facility (HTTF) at the Naval Air Warfare Center, Trenton, NJ. This paper employs classification trees to analyse vibration signatures produced in the HTTF, identifying those characteristics which distinguish "normal" signatures from signatures produced by known faulted parts. These trees are quite successful in separating the two types of signatures and achieve small misclassification rates for HTTF data. They are also applied to vibration data collected from an operational aircraft; assuming the tail gearbox in the operational aircraft has no faults, the trees derived from the HTTF produce a high proportion of false alarms.

14. SUBJECT TERMS
Classification trees, Clustering, Vibration analysis

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

Classification Trees and the Analysis
of Helicopter Vibration Data

Harold Larson

Naval Postgraduate School

Abstract

Health and Usage Monitoring Systems are receiving a great deal of interest, in an attempt to increase the safety and operational readiness of helicopters, and to minimize maintenance costs. These systems monitor (and can record) various flight parameters, pilot conversations, engine exhaust debris, metallic chip detector levels in the lubrication system, rotor track and balance, as well as vibration levels at selected locations throughout the airframe and the power drive system. Vibration levels are currently being observed on two operational SH-60B helicopters and on an H-60 power drive system installed in the Helicopter Transmission Test Facility (HTTF) at the Naval Air Warfare Center, Trenton, NJ. This paper employs classification trees to analyse vibration signatures produced in the HTTF, identifying those characteristics which distinguish "normal" signatures from signatures produced by known faulted parts. These trees are quite successful in separating the two types of signatures and achieve small misclassification rates for HTTF data. They are also applied to vibration data collected from an operational aircraft; assuming the tail gearbox in the operational aircraft has no faults, the trees derived from the HTTF produce a high proportion of false alarms.

Introduction and Background

Rotary wing aircraft lack the inherent stability enjoyed by many fixed wing aircraft if power is suddenly lost. Thus ensuring the integrity of the power system, and its correct employment during flight, is of paramount importance for the safety of the pilot(s), the aircraft and its cargo. Current maintenance policies for helicopters are time-based, requiring the replacement of individual parts after a fixed number of flight hours. This type of policy is designed to be very conservative and can cause removal of good parts well before the end of their useful lifetimes. Even with this approach, a large number of incidents leading to loss of aircraft still occur.

To increase the safety of helicopter flight there has been considerable recent interest in the employment of special monitoring devices on board the aircraft; these typically incorporate advanced mechanical diagnostic technologies for monitoring the propulsion and power drive systems of the aircraft. These are generally called Health and Usage Monitoring Systems (HUMS) and have the ultimate goal of providing real time warnings well in advance of failure-causing incidents. It is also hoped that HUMS may provide the opportunity to employ on-condition maintenance procedures, allowing various parts of the aircraft to remain safely in service longer than durations allowed by fixed flight-time replacement schedules.

The US Navy currently has two different prototype HUMS implementations in place. One of these is installed on an SH-60B at Helicopter Anti-Submarine Squadron, Light-41 (HSL-41) at Naval Air Station (NAS), North Island, California. This HUMS is the result of a Cooperative Research and Development Agreement (CRADA) between the Navy, Smiths Industries, and Chadwick-Helmuth Company, Inc., and uses off-the-shelf equipment provided at no cost to the Navy. Its capabilities include on-board rotor track and balance, engine exhaust debris monitoring, rotor and engine health monitoring, as well as vibration measurement recording; it also includes a crash-survivable combined voice and data recorder. The aircraft on which the HUMS is installed is employed in the same way as the other aircraft in its squadron; several hundred hours of flight-recorded data have been collected. These flight-recorded data are the beginning of a data base library for a helicopter in working order performing the various tasks typical of an HSL squadron aircraft.

The second Navy effort is called the Helicopter Integrated Diagnostic System (HIDS) (Emmerling [2]). This system has been installed on an operational SH-60B aircraft at Naval Air Warfare Center, Aircraft Division, Patuxent River, Maryland. It monitors aircraft usage, condition of the engines, drive shafts, and gearboxes, the chip detectors and performs rotor track and balance. In addition it is capable of the simultaneous recording of up to 32 digital time series channels (from accelerometers and tachometers) at 100,000 samples per second. This high sampling rate allows fine resolution investigation of vibration signals in both the time and frequency domains. In addition to this operational aircraft system, Naval Air Warfare Center, Aircraft Division, Trenton, New Jersey (NAVAIRWARCENACDIVTRENTON) has a test bed (Helicopter Transmission Test Facility, HTTF) with a a full scale SH-60 power drive system, instrumented with the same HIDS. The same vibration sensors are mounted in the same way at the same locations on both the operational aircraft and on the power drive system at the HTTF.

Vibration monitoring

A major goal of the HIDS effort is the establishment of a library of vibration signals. Baseline data is gathered by running the system with all parts known to be in good working order; in the HTTF, as opposed to the operational aircraft, it is also possible to gather vibration signals with faulty parts installed. Some of these faulty parts have been gathered from fleet rejections (where inspection uncovered a faulted part); man-made alterations to good parts (by deliberately cutting slits into gears, chipping a gear tooth, spalling off a piece of a bearing or a bearing-race, *etc.*)

can also be employed. The resulting changes (from the unfaulted baseline cases) in the vibration signals can then be observed; these are expected to lead to algorithms which can be employed on in-flight aircraft with the HIDS system to provide early warning signals of possible problems indicated by recorded vibration levels.

With 100,000 data values simultaneously collected from as many as 32 sensors each second (actually only 26 sensors are currently used), a very large number of possibilities exist for trying to differentiate between the vibration signal envelope of a healthy system versus the vibration signal envelopes of a system which has one or more faulty parts. NAVAIRWARCENACDIVTRENTON has contracted B. F. Goodrich Aerospace-Technology Integration (BFG-TI) to develop vibration diagnostic algorithms from observed HTTF data.

The SH-60 power drive system contains 9 major components: port and starboard engines, port and starboard input modules (deliver power from respective engines), port and starboard accessory drives (receive power from corresponding input modules), main gearbox (receives power from both input modules), intermediate gearbox (receives power from the main gearbox), and tail gearbox (receives power from the intermediate gearbox). Two or more sensors (accelerometers, tachometers) are attached to the drive system housing on or close to each of these major components (the major vibration signals received by a given sensor are presumably from the component it is near).

An HTTF (or aircraft) data acquisition consists of recording the data from the installed sensors over a period of time, typically 4 to 10 seconds (each sensor providing 100,000 values per second). The raw sensor data recorded from each acquisition is archived and can then be examined in both the time and frequency domains. Various measures (called "indicators") can be computed for each data acquisition; the values of these indicators are then to be used to decide whether the data acquisition came from a healthy drive system or from one which may have some problem(s).

Any system that issues warnings about the physical state of the item being monitored (whether it is a helicopter or a human) can commit two different errors: a false alarm (erroneously indicating something is wrong) or a false positive (erroneously indicating nothing is wrong). Each of these errors has its own costs. A false alarm causes unnecessary aircraft downtime (and associated inspection costs). The cost of a false positive depends on what is actually wrong but not detected; the range for this cost can be quite large and may include loss of life, as well the aircraft.

Any complex system may continue working satisfactorily with some level of degradation of specific parts; thus the precise definition of a faulted part plays a role in

3

how successful any diagnostic program may be. An ideal diagnostic system should not only give a warning that a problem exists, but should also provide information pointing at the specific cause of the problem. This allows the possibility that the system correctly identifies that something is wrong, but incorrectly identifies what is wrong. Of course, it is also possible that two or more separate problems exist in the system; ideally all problems should be identified simultaneously.

Typical indicators that can be computed from an HTTF data acquisition can include such things as

- Root Mean Square of the vibration amplitude (at selected frequencies).
- Peak to Peak values, differences between the maximum and minimum amplitudes of a given frequency.
- Skewness and kurtosis of various measures in the time and frequency domains.
- Energy level of strong tones in the spectrum.
- Energy level remaining after removal of strong tones.

The exact definitions of the indicators computed under contract by BFG-TI are proprietary. Since the same problem in the power train might be isolated by any of several different indicators, the initial tendency is to suggest and compute more, not fewer, indicators for a given acquisition. Thus, it appears that the indicators computed by BFG-TI include some redundancy.

Classification trees with HTTF data

Many different classification techniques can be applied to define diagnostic algorithms using computed indicators from a data acquisition. This paper discusses the use of classification trees to isolate different types of faults, using BFG-TI computed indicators from HTTF data acquisitions; it builds on the work of Rovenstine [3].

For his Master of Science thesis in Operations Research at the Naval Postgraduate School, Rovenstine [3] used classification tree methodology in analysing HTTF data provided by NAVAIRWARCENACDIVTRENTON. For this work, he was given the indicators computed by BFG-TI from 640 HTTF data acquisitions made between December 1, 1994 and January 3, 1997; from the data supplied, he extracted the indicators computed from the two accelerometers attached to the housing for the intermediate gearbox. The housing for this gearbox consists of two pieces separated by a gasket to accomodate the bend in the aircraft tail between the main gearbox and the tail rotor. One of the sensors is mounted on the main gearbox side of the gasket and the other is on the tail rotor side; these two sensors are expected to provide information on the health of this intermediate gearbox.

4

Part of the research effort at NAVAIRWARCENACDIVTRENTON is concerned with understanding how initial minor defects in parts may propagate into major faults of concern for flight safety. In support of this effort, an electronic discharge machine (EDM) was used to create a notch in the root of a tooth on the intermediate gearbox input pinion (the gear which brings power from the main gearbox). This faulted piece with one EDM notch was used for 186 of the 640 acquisitions provided to Rovenstine. This effort was actually two-pronged: First, it was desired to see whether the indicators for the intermediate gearbox could differentiate these acquisitions from the 396 baseline cases (no known faults in the intermediate gearbox) included in the 640 acquisitions. Additionally, it was of interest to see whether this EDM notch would propagate into something more serious under the pressure of operational use. This single notch did not propagate into a serious fault over the runs made, so two additional EDM notches were cut into this input pinion and an additional 36 of the 640 acquisitions were made with 3 EDM notches on this input pinion. (This effort was successful; a root bending fatigue crack was propagated in the pinion.) The remaining 22 acquisitions employed an intermediate gearbox input pinion which had one-third of a tooth purposely removed.

A total of 38 BFG-TI indicators are available for each intermediate gearbox sensor for each acquisition, giving 76 available variables for distinguishing the faulted intermediate gearbox cases from the baseline cases, and to distinguish between the 3 different types of faults. Classification trees (Breiman [1]) are appealing as a technique for accomplishing this. This procedure is nonparametric in nature and simple to understand. Like other classification techniques, classification trees require a "training set" of data for building the tree. So long as the data that is used to grow the tree is representative of the operational data to which it is applied, the predictions made by the tree should be accurate and simpler to interpret than those provided by most other procedures.

In applying this technique, the available data (640 cases, 76 observed variables for each) are used recursively to find the best binary splits for the observed variables, thus growing a tree. Briefly, this is done as follows: 396 of the cases are known to be baseline (no fault in the intermediate gearbox) and the remaining 244 are known to be faulted (combining the 3 known types of intermediate gearbox faults). All 640 data values originally reside in a single node (called the root node although it is at the top of the resulting tree). Of these the fraction $396/640 = .61875$ are unfaulted and the remainder are faulted.

First this root node is split into two child nodes. To accomplish this, each of

the 76 variables is examined individually; for each individual variable, the 640 cases could be split into a left child node and a right child node, based on the value of this single variable. Thus, the left and right child nodes could be defined for every possible split point of variable 1. This procedure is applied in turn to each of the 76 variables; that variable, and the split point for that variable, is chosen to produce the "purest" child nodes. Purity of a node is measured by its deviance, which is -2 times the maximized Bernoulli likelihood function (for this situation of fault versus no fault in the intermediate gearbox; with 3 or more classes the natural generalization to the multinomial likelihood is used).

To illlustrate the deviance definition, and its use, suppose the root node contains 640 cases, of which 396 have no fault (success) and the remaining 244 have some fault (failure). If these values were the result of 640 Bernoulli trials which produced 396 successes, the estimated probability of success is 396/640 and the estimated probability of failure is 244/640. The Bernoulli likelihood function, maximized by replacing the probabilities of success and failure by their maximum likelihood estimates (just given), is

$$L = \left(\frac{396}{640}\right)^{396} \left(\frac{244}{640}\right)^{244}.$$

The deviance for the root node then is $D_R = -2\ln L = -2[396\ln 396 + 244\ln 244 - 640\ln 640] = 850.78$. Suppose a variable named IR5.2 is smaller than 12.05 for 228 of these 640 cases, and larger than 12.05 for the remaining 412 cases; the 228 cases go to the left child node and the 412 cases go to the right child node. Let us also suppose that 227 of the cases in the left child are success (one is failure) and that 169 of the cases in the right child node are success (so 243 are failures). Note that the two child nodes account fully for all successes and failures from the root node: $227 + 169 = 396$, $1 + 243 = 244$.

As above, we can define the maximized Bernoulli likelihood function for *each* of these two child nodes, from this get the left and right child deviances and sum them together. This gives $D_l + D_r = 12.85 + 557.79 = 570.64$, a one-third reduction from the deviance in the root node. CART looks at all other possible ways to split the cases from the root node into a left and right child according to the values of IR5.2 and computes $D_l + D_r$, the sum of the child deviances. It does this for all other variables as well. Granted that the split described above, based on IR5.2 being smaller than 12.05, gives the smallest sum for the two child deviances (the biggest reduction possible from the root deviance), CART selects these as the first split in the tree. This procedure is then repeated for each of the two child nodes, again repeated for each of these, and

a classification tree is produced,

Figure 1 should help clarify this discussion. This is a truncated tree using the 640 acquisitions discussed above, each of which is classified as "no fault" (labelled no in the figure) or "fault" (labelled fault). The root node is at the top, represented by the ellipse labelled no; the label for a given node is determined by the majority case present (396 of the 640 cases were no fault). The misclassification rate is printed below the node (244/640 for this root node). Recall that there are 76 computed indicators available with these data; one of these is called IR5.2. Of all the possibilities, across all 76 variables, the "purest" child nodes are given by sending those acquisitions with IR5.2 < 12.05 to the left, with the remaining acquisitions sent to the right child node. The variable, and its values, chosen for the split is used to label the lines drawn between the parent and child nodes. Even though IR5.2 is in theory "continuous" it can take on no more than 640 different values for this data set. The picture of the CART tree uses the midpoint between two successive (ranked) values of the variable for labelling the lines between nodes. (In this case, the ranked set of values for IR5.2 includes 12.00 and 12.10 and no values in between, so the label used is 12.05.)
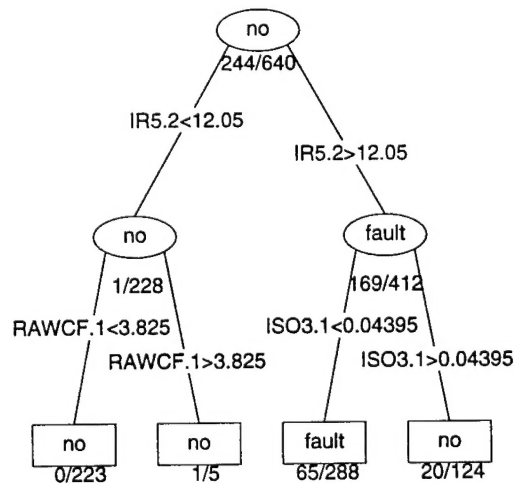

Figure 1. Truncated tree identifying unfaulted acquisitions (labelled no) and faulted acquisitions (labelled fault).

Note that 228 cases go to the left, and the remaining 412 go to the right. The left child node is labelled no because the (great) majority of the 228 acquisitions are "no fault"; in fact, since the misclassification rate below this node is 1/228, we know that 227 of these cases are "no fault" and the remaining one is "fault". The right child node is labelled fault because the majority of these cases are "fault"; from the misclassification rate below the node (169/412) we see immediately that there are 169 "no fault" cases in this node and 412−169=243 cases are "fault".

7

The same procedure is now repeated for each of these two child nodes. The left child node contains 228 cases (and all 76 variables). Each variable again is examined for splitting this node into two parts based on the sum of the two child deviances produced. That variable, and its value, which leads to the greatest reduction in the sum of the two resulting child deviances is selected; for these data, the left child node is split according to whether or not RAWCF.1 < 3.825. The right child node is examined in the same way, leading to the split on the value for ISO3.1. This produces the (purposely truncated) tree with 4 terminal nodes (rectangles) called "leaves", given in Figure 1. Note that with this small tree, 223+4+104=331 of the "no fault" acquisitions have been correctly identified (the remaining 65 were false alarms, misclassified as "fault"); 223 of the "fault" acquisitions have been correctly identified, while the remaining 21 were false negatives, misclassified as "no fault". For this tree, or any other, the final overall misclassification rate is easily determined by examining the leaves (rectangles); the overall rate is given by the ratio of the sum of the numerators to the sum of the denominators for the fractions below the leaves. Thus, for the tree in Figure 1, the overall misclassification rate is $(0+1+65+20)/(223+5+288+124)=86/640$.

Numbering the root node 1, the first two child nodes 2 and 3, and the final four nodes 4, 5, 6, 7 (from left to right), it is apparent that there is nothing to gain in trying to split node 4 (it is "pure", contains only "no fault" acquisitions, its deviance is 0), and very little to gain from splitting node 5. Further splitting would be useful though, for nodes 6 and 7, giving better separation of the two types of acquisitions for these 412 (288+124) cases. In practice, this would of course be done (and has been done; see [3] and discussion below).

Classification trees, like Figure 1, produce a natural multi-dimensional extension of single-variable exceedance limits for distinguishing the two types of acquisitions. That is, by tracing the left-most side of the tree, an acquisition with IR5.2<12.05 and RAWCF.1<3.825 is certain to be "no fault", for this set of data. An acquisition with IR5.2>12.05 and ISO3.1<0.04395 is likely to be "fault"; in fact the estimated probability it is not "fault" is 65/228=.285. The tree structure will typically suggest that the values of two or more of the variables observed should be used to distinguish "fault" from "no fault", rather than the value of a single variable.

With a large number of "continuous" variables available for building the classification tree, as in this case, one could continue splitting until every node is pure. (This of course requires that there does not exist two or more cases which have identical values for all 76 classification variables, some of which are fault and some of which are no fault.) It is usual to let CART continue growing the tree until all "relevant" struc-

8

ture is found. This typically means growing a large tree that may "overfit" the data, including small details driven by the particular data set observed. This overlarge tree may or may not be successful in correctly classifying new data acquisitions. Cross validation is typically used in choosing the correct size for a tree; this procedure is discussed and illustrated below.

To summarize some facts and conventions about classification trees:

- Nodes are labelled with the identity of the majority of cases included.
- The fraction below a node gives the misclassification rate for the node (number of acquisitions included that are *not* of the class used for the label).
- The variable used for splitting the cases in the parent into the two child nodes, and its values, are printed on the lines connecting the parent node to its children.
- The sum of the denominators of the misclassification rates below the two children is equal to the denominator of the misclassification rate of the parent (all cases included in the parent must be accounted for in its children).
- The sum of the numerators of the misclassification rates below the two children is smaller than the numerator of the misclassification rate of the parent (because any split must lead to "purer" nodes).
- The overall misclassification rate for the tree is given by the ratio of the sum of the numerators to the sum of the denominators for the fractions below the terminal leaves (rectangles at the bottom of the tree).

Rovenstine [3] used these 640 HTTF acquisitions to grow two classification trees; his model 2 employed two categories for the acquisitions, faulted (244 cases) versus no fault (396 cases). His model 1 employed 4 categories for the acquisitions, no fault (396), single EDM notch (186), three EDM notches (36) and tooth failure (22). He used cross validation to choose the best tree size for both models and additionally employed a heuristic method to search for the best trees of the given sizes. Both of these were quite successful, giving useful trees for classifications of the two kinds. His model 1 tree included 11 terminal leaves, with 16 missed faults and 4 false alarms (as well as 3 faults misclassified) for an overall misclassification rate of $23/640=.0359$. His model 2 tree included 12 terminal leaves, with 10 missed faults and 7 false alarms, for a misclassification rate of $17/640=.0266$.

After Rovenstine had finished his thesis, a more careful look at the 640 HTTF acquisitions available revealed further detail about the 396 baseline acquisitions provided. Of these, only 32 consisted of acquisitions for which there were no known faulted parts anywhere in the power drive system run on the HTTF (true or "pure" baseline cases); each of the other 364 acquisitions had known faults in one or more

9

of the other components of the power drive system (but none in the intermediate gearbox). The two sensors Rovenstine used are attached to the housing of the intermediate gearbox, as already noted; since vibrations may find paths through the system, the question arises whether the intermediate gearbox sensors might in fact be able to detect faulted parts in other components of the power drive system.
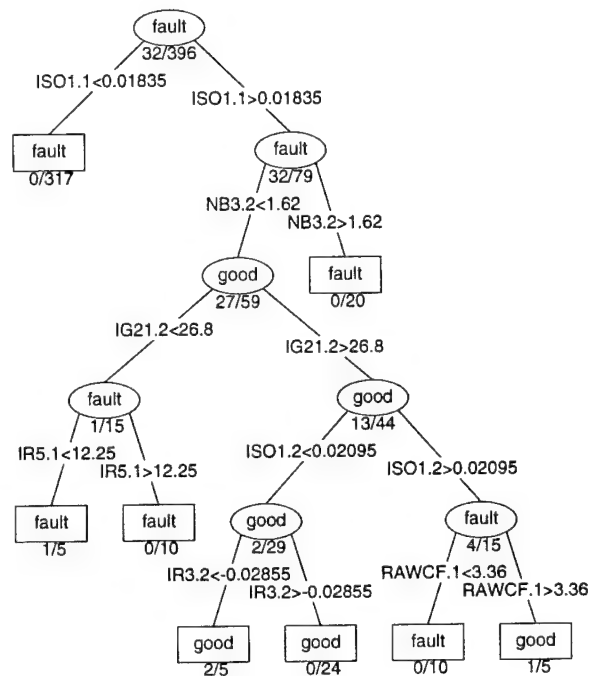

Figure 2. Tree identifying true baseline cases. no signifies no fault in the system, fault indicates one or more known faults somewhere in the system

To investigate this point, the original 396 baseline cases were categorized as no (32 cases) or fault (remaining 364 cases) and a classification tree was grown, using the software defaults. This means that node splitting continues unless the node is pure, or the node contains 5 or fewer cases. The resulting tree is presented in Figure 2.

The first split (from the root node at the top), using the value of ISO1.1 correctly identified 317 of the 364 acquisitions which had faults somewhere in the power drive system; the second split, made on the value of NB3.2 correctly isolated 20 more of the remaining 47. The remaining splits were made on the values of variables IG21.2, IR5.1, ISO1.2, IR3.2 and RAWCF.1. This tree performs very well on the full set of data. Only one no (acquisition with no faulted components in the system) is classified as fault (a false alarm) and three fault acquisitions are classified no (false negatives), for an overall misclassification rate of 4/396, slightly above 1%.

CART can also prune trees by removing nodes, at a cost of increasing the de-

10

viance. It is of interest to examine the change in deviance for different sized trees, resulting from pruning back the size of the tree. Figure 3 presents a plot of the deviance for this tree versus the size of the tree. Note that by far the greatest reduction in deviance comes from having two nodes instead of one (caused by the intial split which correctly recognized 317 of the 364 faulted acquisitions). The second largest change in deviance occurs in going from 4 to 5 nodes. This apparent success in giving a very small misclassification rate may be due to the tree being overly large, overfitting the data.
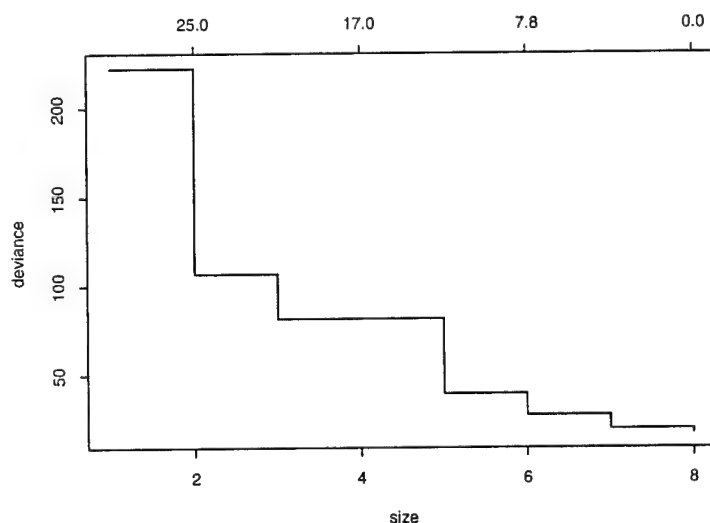

Figure 3. Deviance versus size of tree for Figure 2.

Cross validation can be used to choose the right size for a tree. This procedure uses a portion of the observed acquisitions to grow a tree and then sends the acquisitions not used down the tree to test its predictive ability. Ten-fold cross validation has been found empirically to perform quite well in choosing the "best" size for a tree ([1]), in which the data set is broken (at random) into ten roughly equal sized groups. This gives ten mutually exclusive groups, each of (roughly) 1/10 the size of the original data set. For each of these, the other 90% is used for tree growing, producing a *sequence* of trees increasing in size from the root node to a tree the same size as grown with the full data set. For each tree in the sequence, the cases held out from the tree-growing are run down the tree and the deviance is computed. This is repeated for each of the other nine sets of data, resulting in ten computed deviances for each size tree. These ten deviances are then summed together and plotted versus the tree size; the "best" size is identified by the minimum deviance in this plot. Figure 4 presents a cross validation plot of the tree given in Figure 2.

11
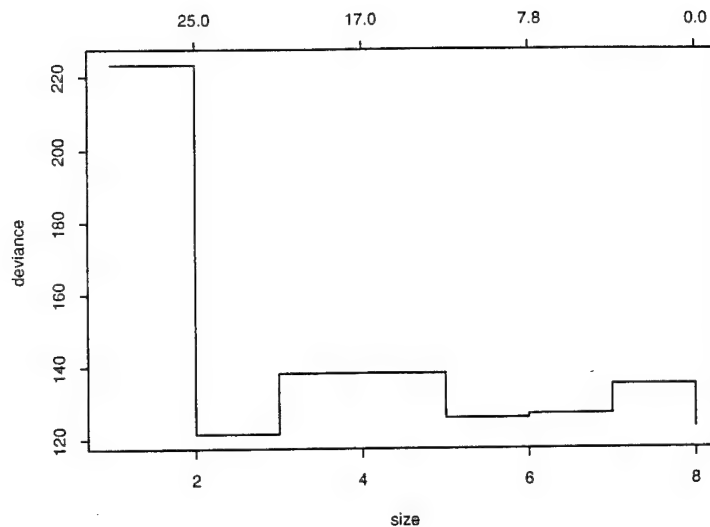
Figure 4. Cross validation plot for tree given in Figure 2.

The two-node tree gives the smallest sum of the deviances through cross validation; going to three or four nodes increases the sum of the deviances, while it decreases again for five nodes. Since the sum of the deviances for the five node tree is almost as small as for the two node tree, one should expect fairly good predictions from a five node tree. This best five-node tree is given in Figure 5. It has a misclassification rate of 7/396, under 2%.
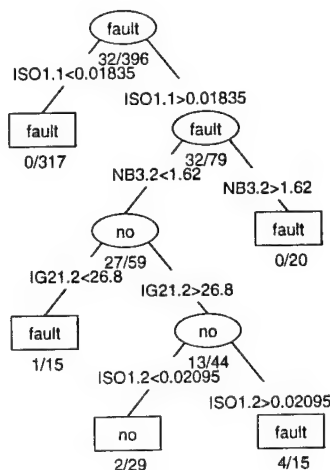
Figure 5. Five-node tree for identifying faults outside the intermediate gearbox.

It appears that the sensors attached to the intermediate gearbox housing have

12

some success in detecting faults in other parts of the system. How accurate are these intermediate gearbox sensor indicators in distinguishing the various faults in other parts of the power drive system? The 364 acquisitions with faults in components other than the intermediate gearbox contained a variety of components with faulted parts; some components were used in the port input module for some of the acquisitions and in the starboard input module for others. These acquisitions with faulted parts were separated into 12 categories (making 13 in total including the true baseline cases). Again the true baseline cases (no fault anywhere) were labelled no, and the acquisitions with one or more faults were numbered from 2 through 13. Table 1 gives an indication of where the faulted components were located. The column labelled Number reports the number of acquisitions with some fault in the indicated component. Thus, all 364 baseline cases actually had a fault in the tail gearbox; 163 cases had a fault in the port input module. The row labelled Total gives the numbers of acquisitions in categories 2 through 13; these total to 364.

Table 1. Locations of faulted components in baseline cases.

Component	Fault Class												Number
	2	3	4	5	6	7	8	9	10	11	12	13	
Tail gearbox	√	√	√	√	√	√	√	√	√	√	√	√	364
Main gearbox	√	√	√										128
Starboard input mod		√	√				√	√	√	√	√	√	228
Port input mod						√	√	√	√	√	√	√	163
Total	33	24	71	42	10	5	29	45	29	9	51	16	

A classification tree was grown using these 396 baseline acquisitions, with 13 different categories. The resulting tree was rather large, with 23 leaves, overfitting the given data (it had a misclassification rate of 45/396=.1136). Cross validation was used to find the best sized tree, as above, resulting in the 16-leaf tree given in Figure 6. The root node is labelled 4, since this class of faults is most frequent (see the totals in Table 1). The first split, on ISO1.1, correctly sends all 71 category 4 acquisitions to the left child node; the majority of those sent to the right child node are class no. All 32 truly unfaulted acquisitions are in this node. The path of the developing tree can be followed in the figure. One leaf is labelled no, with a misclassification rate of 14/42=.3; thus 28 of the 32 true baseline acquisitions are correctly identified in this node (and 14 faults are here as well). The overall misclassification rate is 69/396=.1742; this is surprisingly good for sensors which are not placed to identify faults in these various locations. The appendix contains more information about

the categories used and details about which categories were misclassified (between themselves or as being unfaulted).
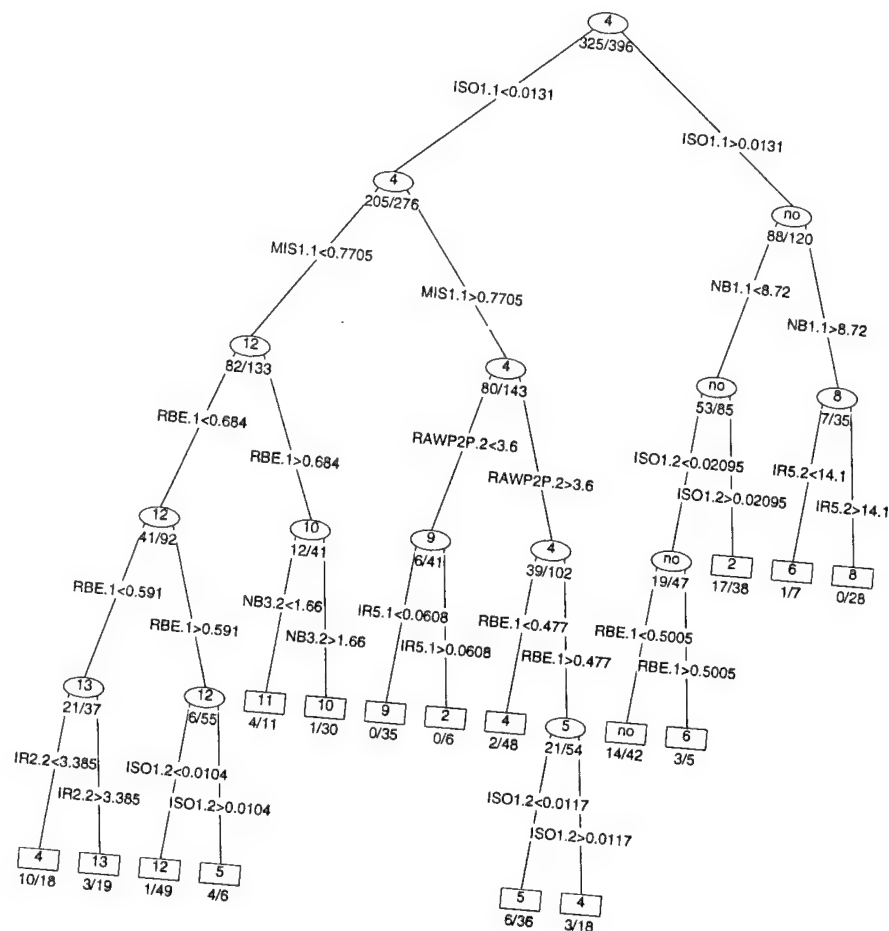
Figure 6. Tree identifying faults outside the intermediate gearbox. no indicates unfaulted acquisitions. 2, 3,..., 13 indicate acquisitions with faults.

The fact that the intermediate gearbox sensors are capable of identifying faults in other components suggests that a classification tree using only the 32 true baseline acquisitions, together with the 244 acquisitions with a fault in the intermediate gearbox, might be different than that grown by Rovenstine. The classification tree using no to identify the true baseline acquisitions and fault to identify the faulted intermediate gearbox acquisitions is given in Figure 7. (Cross validation pruned the tree back to five leaves versus the original six.) This tree has a misclassification rate of 2/276=.0072, with both misclassifications being false alarms (no missed faults). The variables defining the first two splits (RBE.2 and IR2.2) both also occur, in the same order, in Rovenstine's tree, but considerably further down his tree. This tree,

14

with five leaves, is considerably simpler than Rovenstine's model 2 tree; of course, this tree is based on only 276 of his 640 acquisitions.

Figure 7. Tree identifying true baseline (no) versus faults in the intermediate gearbox (fault).

Next a tree was grown using the 32 true baseline acquisitions, again with the 244 faulted intermediate gearbox acquisitions, but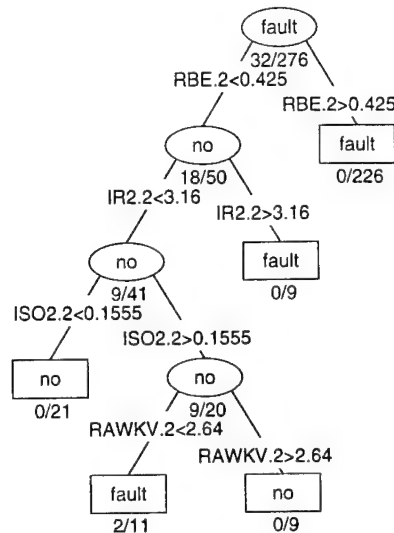 this time identifying the type of fault in the intermediate gearbox. The tree produced had nine leaves with 3 misclassifications. Cross validation suggested pruning this tree back to five leaves, as given in Figure 8. It has ten false alarms and one misclassified fault (one edm was misclassifed as edmthree), for a misclassification rate of 11/276, about 4%. The false positives are all tooth failures.

The trees presented in Figures 7 and 8 used all of the faulted acquisitions for the intermediate gearbox. The faults labelled edm were introduced with two goals in mind. First and foremost, it was of interest to see if they could be detected. Secondly, serious cracks in this pinion must start in some way; it was thought that the introduced slit might propagate into a serious crack in the input pinion (this did not occur with the single notch). The HTTF true baseline acquisitions (and many of the other acquisitions) were typically acquired in groups of about six. The first acquisition in one of these groups of six generally began soon after the the engines were started, with cold oil and moderate torque on both the main rotor shaft (about 100 ft-lb) and the tail rotor shaft (about 400 ft-lb). Then the torque was quickly increased for both shafts (to about 325 ft-lb main, 2000 ft-lb tail) for the second acquisition; the third through sixth acquisitions were then made at torques stepped back down fairly evenly to the beginning values for the sixth and final acquisition.
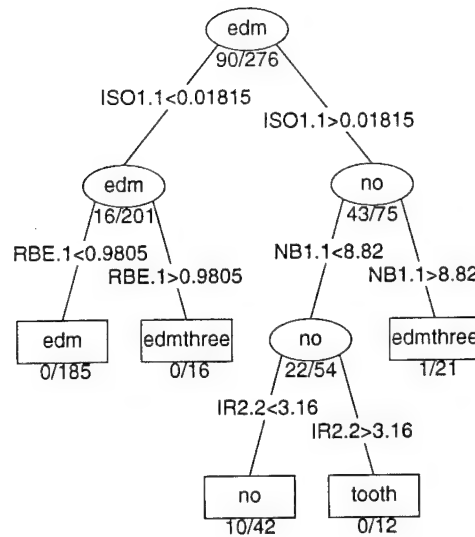
Figure 8. Tree identifying true baseline acquisitions (no) and
fault types (edm, edmthree, tooth)

When trying to propagate the edm slit, though, the pattern of torques applied was frequently quite different (and acquisitions were frequently made in larger groups than just six). One sequence of 44 acquisitions with the single edm slit in the intermediate gearbox input pinion consistently used either 2000 ft-lb torque for the tail rotor shaft (12 acquisitions) or 2400 ft-lb (32 acquisitions). This type of pattern was also used with the edmthree faults (the tail rotor torque was constant at 2350 ft-lb for all 36 acquisitions).

Recall that the great majority of the faulted intermediate gearboxes had the edm fault (186 cases out of a total of 244 faulted gearbox acquisitions). Since the edm acquisitions were made in quite different ways, two additional classification trees were grown. The first of these used the 44 edm high tail rotor torque cases mentioned above, whose conditions were similar to those used for the edmthree acquisitions. The resulting tree, after cross validation, is presented on the left in Figure 9. An earlier sequence of 40 edm acquisitions, which varied the tail rotor torque more like the baseline cases, was used with the true baseline acquisitions (and the edmthree and tooth acquisitions) to grow another tree; cross validation produced the tree presented on the right in Figure 9.

Both of these trees are quite similar to Figure 8; the high tail torque tree on the left uses NB1.1 to split off the edm and edmthree acquisitions, then separating them by the value of RBE1.1. The tree grown with the more usual tail rotor torque patterns (on the right) immediately correctly splits off the edmthree acquisitions using NB1.2 and then isolates the edm acquisitions using ISO1.1. Although different sequences

16

of variables are used on the two trees, they arrive at identical results. Both trees correctly identify the edm and edmthree faults, and incorrectly misclassify the same 10 tooth faults as being no. The different patterns of torque application do not appear to affect the predictive ability of the classification trees.
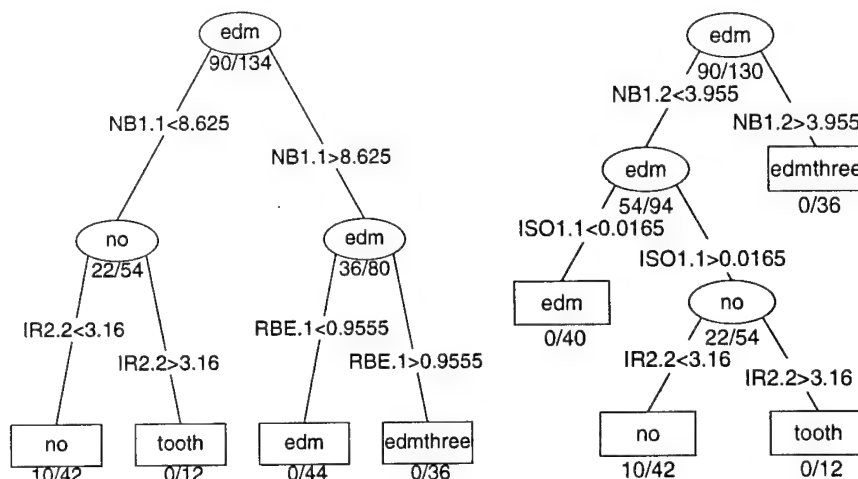
Figure 9. Tree using true baseline acquisitions and different edm acquisitions.
Left tree: High tail torque.
Right tree: Standard tail torques.

The data provided to Rovenstine also included 113 acquisitions from the operational SH-60B helicopter at PAX River. He extracted the data for the intermediate gearbox sensors from these acquisitions and found that 87 of these acquisitions had no real data in them; every variable had value 0 for every acquisition. The remaining 26 acquisitions contained apparently valid data, allowing him to run these acquisitions through his two trees to see how they would be classified. Presumably all 26 of these acquisitions should be classified as no fault since they came from the operating aircraft. Using his model 1 tree (where the three different types of fault were distinguished), he found his tree classified two acquisitions as being edm and four were classified as edmthree, resulting in six false alarms. His model 2 tree (where acquisitions were simply classified as fault versus no fault) produced only two false alarms.

These same operational aircraft acquisitions have also been classified by the trees discussed here, using only the 32 true baseline acquisitions; the results differ from those given by Rovenstine's trees. The tree in Figure 7, where all acquisitions were classified as either no fault or fault, produces 17 false alarms, with 9 acquisitions

17

classified as no. The tree in Figure 8, and the two trees presented in Figure 9, provided better predictions; each of these produced eleven false alarms. Of these, the tree in Figure 8 and the tree on the right in Figure 9 each classified (the same) ten aircraft acquisitions as tooth faults and (the same) one as an edm fault. The left-hand tree in Figure 9 classified all eleven of these aircraft acquisitions as tooth faults.

For some reason, the operational aircraft acquisitions from the intermediate gearbox sensors appear to be quite different than the 32 HTTF baseline cases, leading to the high false alarm rates mentioned above. The major contributor to these differences may very well be the difference between a helicopter power drive system bolted into a concrete test bed and one operating in a flying aircraft. It would seem natural that vibrations will be dampened in a system which is bolted into concrete versus one which is attached to an airframe. This in turn may cause trees grown from HTTF baseline data to produce overly "conservative" breakpoints. Of course, it is also possible (and hopefully unlikely) that there may be some unknown fault in the aircraft's intermediate gearbox.

Conclusions

Classification trees are quite capable of distinguishing faulted and unfaulted HTTF acquisitions, using indicators computed from the raw data observed for the intermediate gearbox sensors. They identify the indicator(s) which are important in identifying faults, and in distinguishing between different types of faults, with no underlying distributional assumptions. Unlike regression-type procedures, the path followed through a classification tree to place a data acquisition into a terminal leaf is determined by simple splits on values of individual indicators. Such paths are easy to understand and are the natural multivariable analog of single-indicator exceedance limits for distinguishing between possible classes to which an acquisition may belong.

The intermediate gearbox sensors are able to identify faults in other parts of the system as well, using the HTTF acquisitions. The trees grown with pure baseline acquisitions (no known fault in any part of the system) can be simple and have a small tree misclassification rate. Those trees which include baseline acquisitions with no fault in the intermediate gearbox (but do have faults elsewhere in the system) are bigger and have a larger tree misclassification rate. This effect is probably due to the "noise" introduced to the intermediate gearbox sensors by the extraneous faults. These latter, larger trees did better in classifying the 26 operational aircraft acquisitions examined, assuming there are no faulty parts in the aircraft's power train system.

18

References

[1] Breiman, L., Friedman, J., Olshen, R., Stone, C. *Classification and Regression Trees*, Wadsworth, Inc., Belmont, California, 1984.

[2] Emmerling, C. *Helicopter Integrated Diagnostic System (HIDS)*, NAVAIRWAR CENACDIV-TRENTON-PPE-271, Naval Air Warfare Center, Aircraft Division, Trenton, New Jersey, November, 1995.

[3] Rovenstine, M. *Classification Analysis of Vibration Data from SH-60B Helicopter Transmission Test Facility*, MS Thesis in Operations Research, Naval Postgraduate School, March, 1997.

Appendix

This appendix discusses the classification of the HTTF baseline acquisitions into 13 classes, separating the true baseline cases (no fault anywhere in the system) from those cases in which there were known faults in other components. It identifies the acquisitions whose identities were correctly indicated by CART, as well as the correct identities of those acquisitions which were misclassifed by CART.

As mentioned earlier, there are 9 major components in the H-60 power system: port and starboard engines, port and starboard input modules, port and starboard accessory drives, main gearbox, intermediate gearbox, and tail gearbox. Thirty-two of the HTTF data acquisitions were made with no known faults; the other 364 baseline cases had known faults in one or more of the tail gearbox, main gearbox, starboard or port input modules. No acquisitions were made with known faults in the engines or the accessory drives.

Across the 396 acquisitions available, seven different input modules were employed (2 with no faults, 3 only with faults and the remaining two both with and without faults). These input modlules were switched in and out of the power drive system, over this time span; for some acquisitions the same two input modules were used, but interchanged between the port and starboard sides. Two accessory drives were used, neither had any faults; they were switched between port and starboard sides at various times. Five different main gearboxes were used; four of these had no faults and the fifth was used both with and without known faults. Two different tail gearboxes were used; one of these had no faults, the other was used both with and without known faults.

no identifies the true baseline cases (no known faults anywhere); the remaining 12 classes had no fault in the intermediate gearbox, but one or more faults elsewhere in the system. Acquisitions were placed in a new class with any change in the confifuration of the system; in particular this includes cases in which exactly the same components were used, but the port and starboard input modules were interchanged. The total number of known simultaneous faults in the system varied from one to three.Table A1 below (this is the same as Table 1 presented earlier) elaborates these definitions, using $\sqrt{}$ to indicate a known faulted part in the indicated component. With the exception of classes 2 and 3 (ten class 2 acquisitions preceded class 3, the remainder followed class 3), the progression of class numbers also indicates actual calendar time as the acquisitions were made. That is, the calendar dates on which class 4 acquisitions were made all preceded the dates on which acquisitions in classes 5 through 12 were made. The sixteen class 12 acquisitions occurred on the latest dates.

Table A1. Locations of faulted components in baseline cases.

Component	Fault Class												Number
	2	3	4	5	6	7	8	9	10	11	12	13	
Tail gearbox	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	364
Main gearbox	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$										128
Starboard input mod		$\sqrt{}$	$\sqrt{}$				$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	228
Port input mod						$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		163
Total	33	24	71	42	10	5	29	45	29	9	51	16	

The tree grown with the 396 baseline acquisitions, classified into groups labelled no and 2 through 13 as described above, had 23 leaves; it had a misclassification rate of 45/396=.1136. Table A2 describes this tree. The columns identify the the 13 different groups; the totals of these columns give the actual numbers of acquisitions in the groups. The penultimate column in this table lists the label on the leaf into which the acquisitions were placed by this tree.

A total of 34 acquisitions were classified no, while 28 of these were actually no, three were 2 and three were 3. None of the leaves was labelled as group 7. (There were only 5 acquisitions truly in group 7.) The final column identifies how many of the 23 leaves contained the various group labels. The "principal diagonal" of this table identifies the correct classifications, while the counts off this diagonal represent misclassifications.

Table A2. Description of 23 leaf tree separating baseline cases.

	Actual class												Classified as	Number of leaves
no	2	3	4	5	6	7	8	9	10	11	12	13		
28	3	3	0	0	0	0	0	0	0	0	0	0	no	2
3	27	5	0	0	0	0	0	0	0	0	0	0	2	3
1	0	16	0	0	0	0	0	0	0	0	0	0	3	2
0	2	0	69	3	0	0	0	1	0	0	1	0	4	3
0	0	0	1	38	0	0	1	3	0	0	1	0	5	3
0	1	0	1	1	10	5	0	0	0	0	0	0	6	3
0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
0	0	0	0	0	0	0	28	0	0	0	0	0	8	1
0	0	0	0	0	0	0	0	35	0	0	0	0	9	1
0	0	0	0	0	0	0	0	1	29	0	0	0	10	1
0	0	0	0	0	0	0	0	4	0	7	0	0	11	1
0	0	0	0	0	0	0	0	0	0	1	48	0	12	1
0	0	0	0	0	0	0	0	1	0	1	1	16	13	2
Totals 32	33	24	71	42	10	5	29	45	29	9	51	16		23

Cross-validation suggested the original tree with 23 leaves should be cut back to 16 leaves (this is the tree presented in Figure 6 earlier). Table A3 describes how the various acquisitions were classified by this tree; it has a misclassification rate of 69/396=.1742. After the tree was pruned back to 16 leaves, label 3 does not occur, and of course there is still no leaf labelled 7. In addition, this pruning results in fewer leaves for classes 1, 2, 5, 6, and 13.

Table A3. Best sized 16-leaf tree.

	Actual class												Classified as	Number of leaves
no	2	3	4	5	6	7	8	9	10	11	12	13		
28	3	11	0	0	0	0	0	0	0	0	0	0	no	1
4	27	13	0	0	0	0	0	0	0	0	0	0	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
0	2	0	69	10	0	0	0	2	0	0	1	0	4	3
0	0	0	2	32	2	2	1	2	0	0	1	0	5	2
0	1	0	0	0	8	3	0	0	0	0	0	0	6	2
0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
0	0	0	0	0	0	0	28	0	0	0	0	0	8	1
0	0	0	0	0	0	0	0	35	0	0	0	0	9	1
0	0	0	0	0	0	0	0	1	29	0	0	0	10	1
0	0	0	0	0	0	0	0	4	0	7	0	0	11	1
0	0	0	0	0	0	0	0	0	0	1	48	0	12	1
0	0	0	0	0	0	0	0	1	0	1	1	16	13	1
Totals 32	33	24	71	42	10	5	29	45	29	9	51	16		16

INITIAL DISTRIBUTION LIST

1. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013)...2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center..2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Therese Bilodeau 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Harold Larson (Code OR/La)..1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Robert Read (Code OR/Re) 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. CAPT Frank Petho (Code NS/Pe) 1
   Dept of National Security Affairs
   Naval Postgraduate School
   Monterey, CA 93943-5000

8. NAWC Aircraft Division...1
   Code PPE, ATTN: M. Mimnaugh
   1440 Parkway Avenue
   Trenton, NJ 08628-3011

9. R.F. Healing, Director 1
   Safety & Survivability, DON
   Rm. 162, Crystal Plaza 5
   2211 Jefferson Davis Hwy.
   Arlington, VA 22244-5107